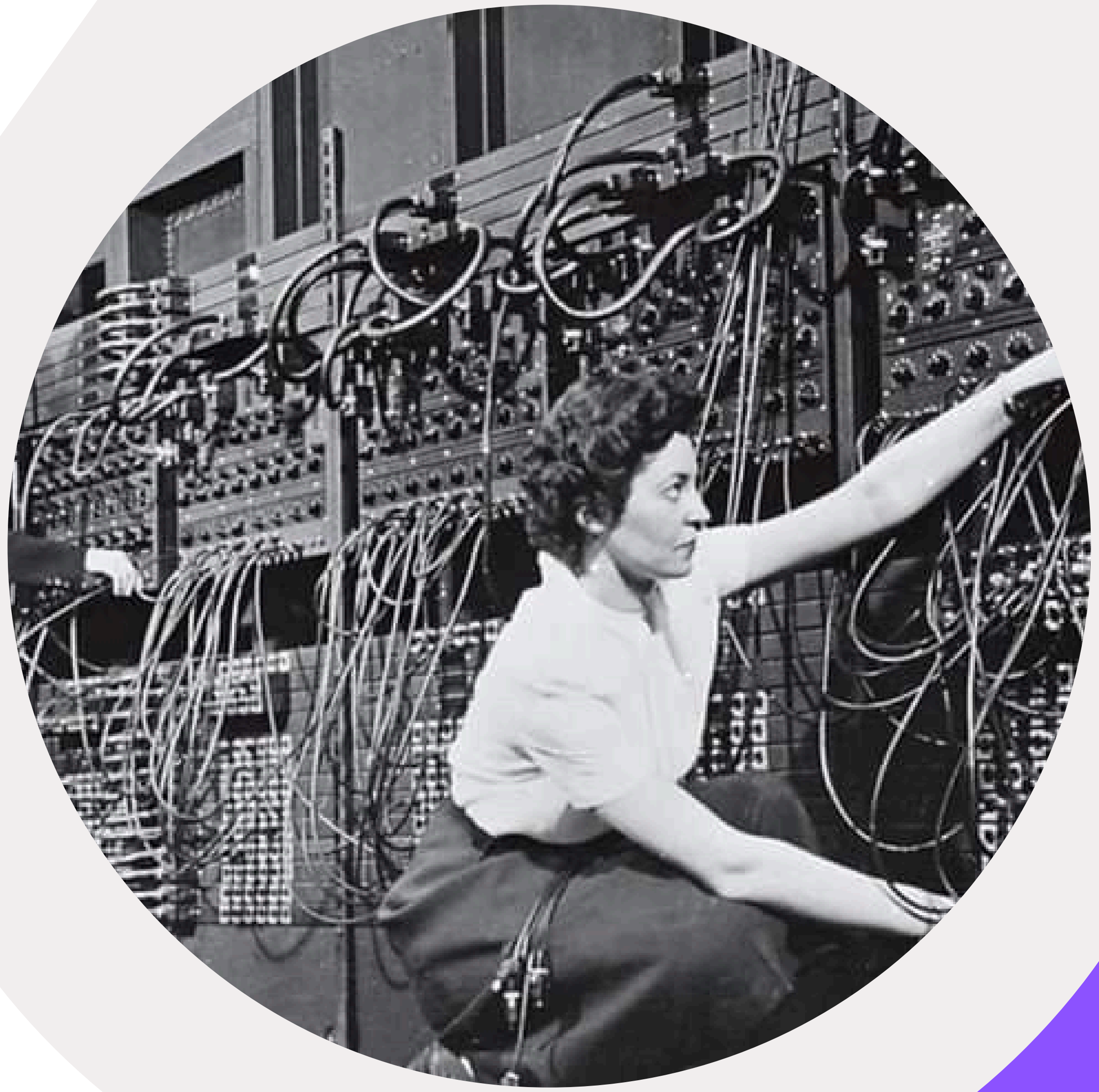
CAN YALÇIN

MACHINE LEARNING

Machine Learning for Climate Prediction: Insights from ClimateWins



Contents

CAN YALÇIN

European Climate Assessment & Dataset

Data Overview

The European Climate Assessment & Dataset (ECAD) project compiles weather data from 18 stations across Europe from the late 1800s to 2022. It includes daily variables like temperature, wind speed, precipitation, snowfall, and global radiation. ECAD's continuous measurements provide insights into regional climate patterns, long-term trends, and extreme weather predictions in mainland Europe, supporting ClimateWins' machine learning initiatives for climate change and planning.

Gündem Sayfasına Geri Dön

CAN YALÇIN

Hypothesis 1: Weather Extremes Prediction

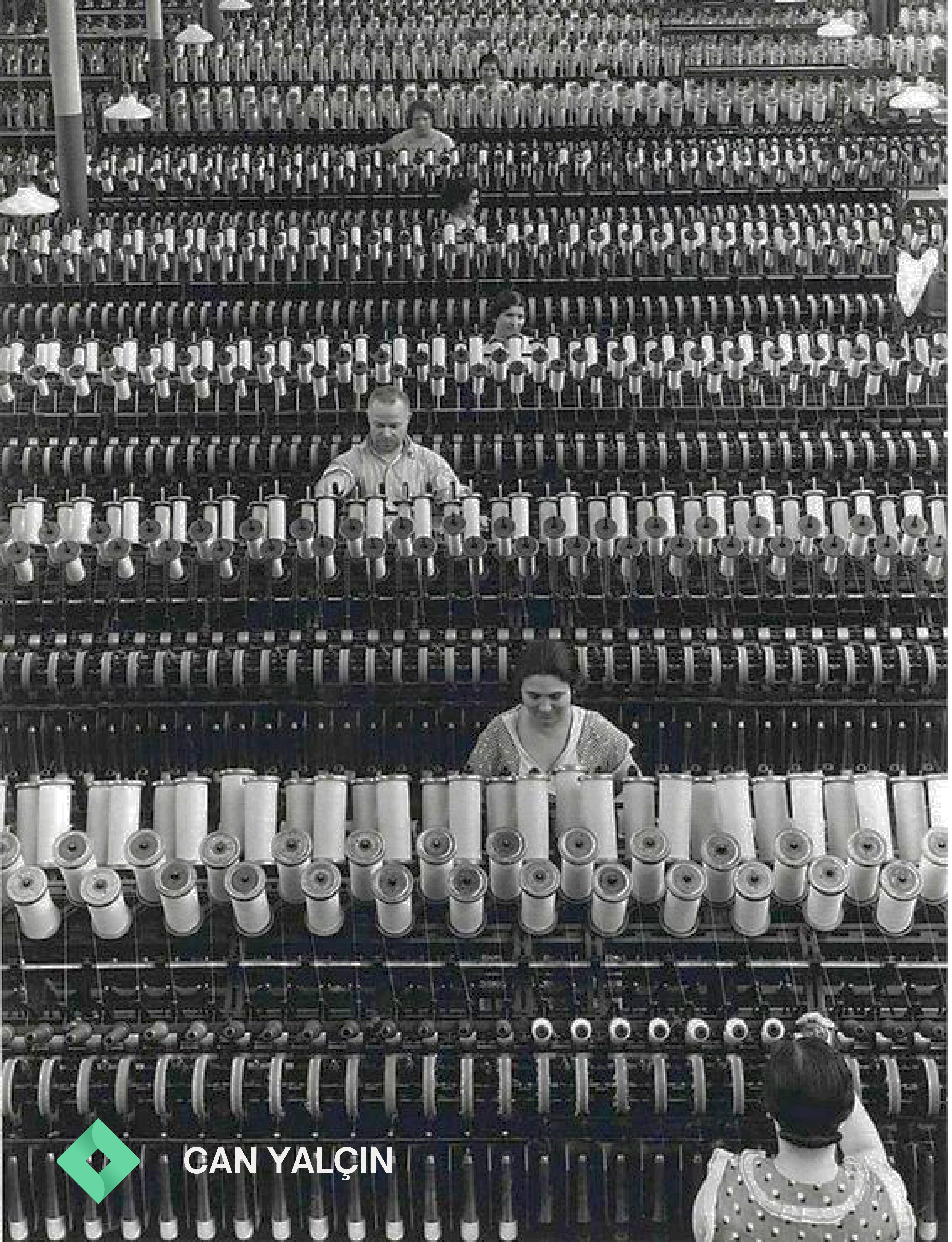
Hypothesis: Machine learning models trained on historical weather data can accurately predict the occurrence and intensity of extreme weather events in mainland Europe.

Gündem Sayfasına Geri Dön

Hypothesis 2: Long-term Climate Trends

- Hypothesis: Analyzing temperature and weather data spanning the past century will reveal significant long-term climate trends indicative of global warming and climate variability in Europe.

Gündem Sayfasına Geri Dön



Hypothesis 3: Model Generalization

- Hypothesis: Machine learning models developed using diverse datasets, including hurricane predictions and global temperature records, can generalize well to predict weather patterns across different regions in Europe.

Gündem Sayfasına Geri Dön

CAN YALÇIN

Data Optimization

Techniques Used

- Gradient Descent: Identifies the local minimum to optimize model performance
- Loss Function: Measures deviation between predicted and actual weather data.
- Scatterplot: Analyzes one year of temperature data over time.
- 3D Visualizations: Plots all weather data for all stations over a year.

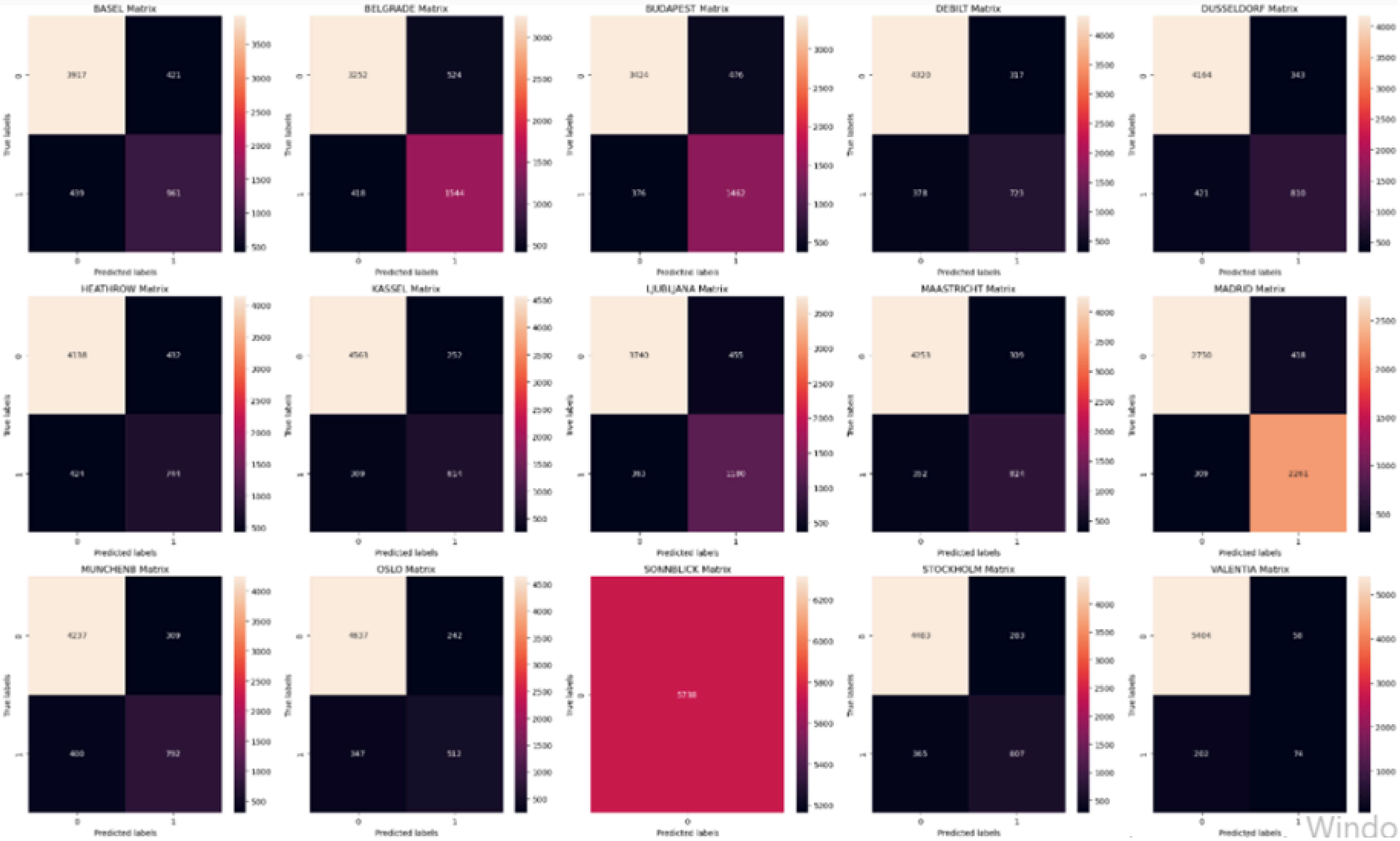
CAN YALÇIN

K-Nearest Neighbor (KNN) Algorithm

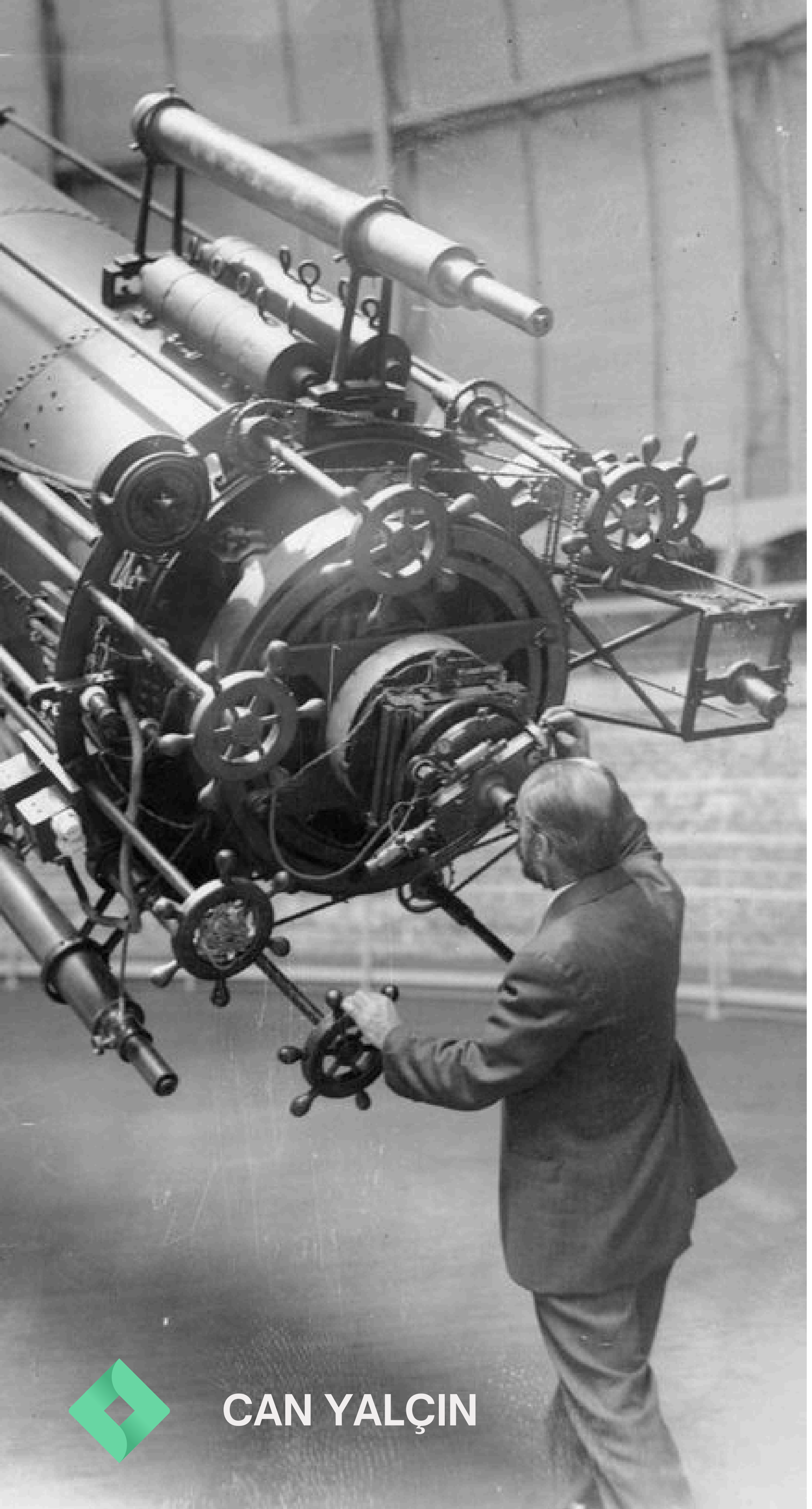
Findings

Gündem Sayfasına Geri Dön

We will use the KNN algorithm to classify and predict climate data. This includes preprocessing the data, splitting it into training and testing sets, and running KNN with various values of k. The model's performance will be evaluated using confusion matrices and accuracy plots to provide insights for ClimateWins.


BASEL Matrix
BELGRADE Matrix
BUDAPEST Matrix
DEBILT Matrix
DUSSELDORF Matrix
HEATHROW Matrix
KASSEL Matrix
LJUBLJANA Matrix
MAASTRICHT Matrix
MADRID Matrix
MUNCHENB Matrix
OSLO Matrix
SONNBLICK Matrix
STOCKHOLM Matrix
VALENTIA Matrix
True labels
Predicted labels

Weather Station	Accurate predictions		False positive	False negative	Accuracy rate
Basel	3917	961	421	439	85%
Belgrade	3252	1544	524	418	84%
Budapest	3424	1462	476	376	85%
Debilt	4320	723	317	378	88%
Desseldorf	4164	810	343	421	87%
Heathrow	4138	744	432	424	85%
Kassel	4563	614	252	309	90%
Ljubljana	3740	1180	455	363	86%
Maastricht	4253	824	309	352	88%
Madrid	2750	2261	418	309	87%
Munchenb	4237	792	309	400	88%
Oslo	4637	512	242	347	90%
Sonnblick	5738	0	0	0	100%
Stockholm	4483	607	283	365	89%
Valentia	5404	74	50	202	96%
				Average	88%



Analysis of Weather Prediction Model Performance

Key Observations

ACCURACY VARIABILITY:

The model shows varying accuracy across different stations, with Sonnblick achieving 100% accuracy in predicting unpleasant weather. However, other stations like Madrid show lower accuracy, especially for pleasant weather.

OVERFITTING CONCERNS:

- The perfect accuracy at Sonnblick suggests possible overfitting, where the model performs well on training data but struggles with new, unseen data.

GENERALIZABILITY ISSUES:

- The model's varying performance indicates it may not generalize well across different weather patterns and locations, highlighting a need for more diverse training data.

Recommendations

REASSESS MODEL ACCURACY:

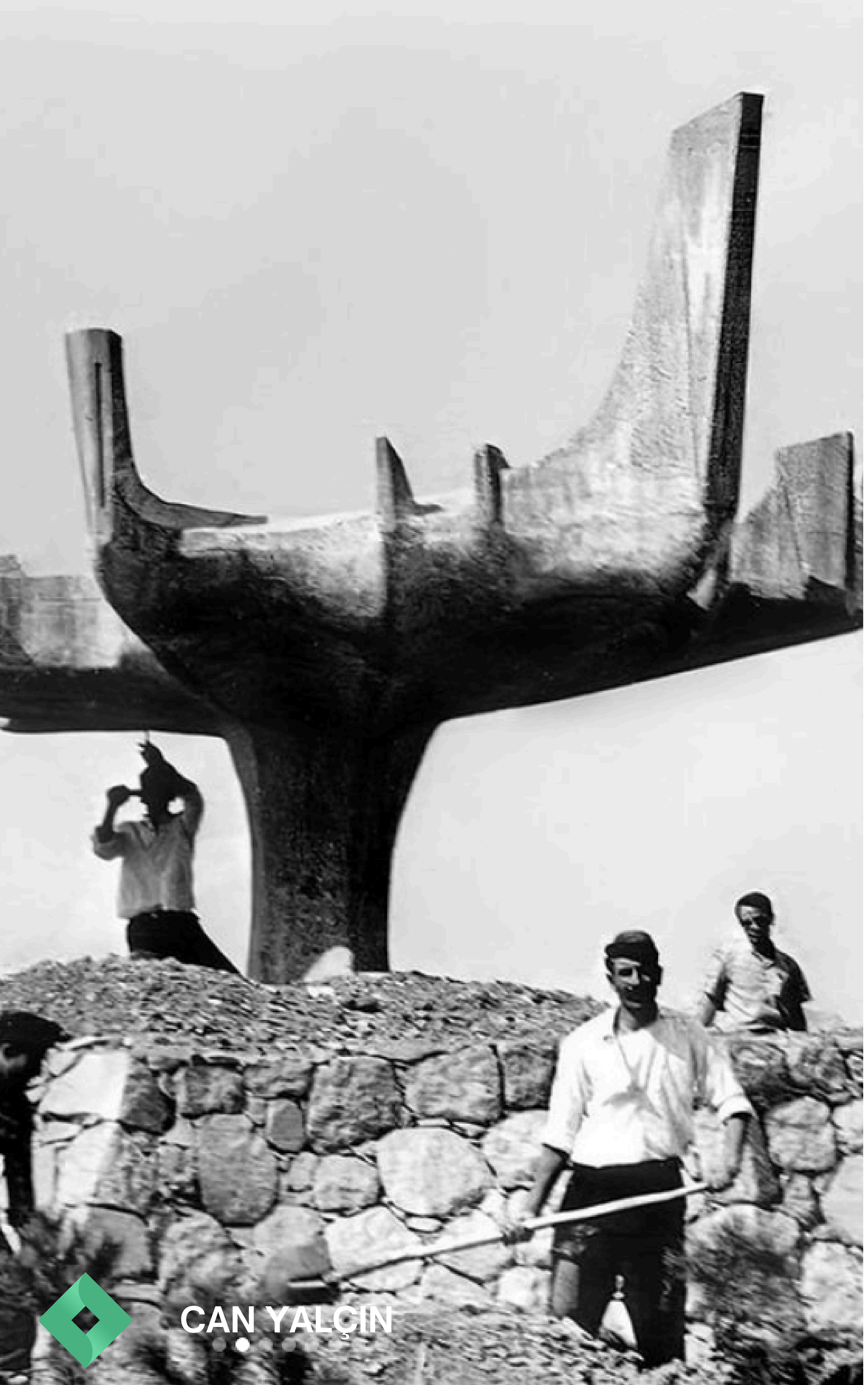
Evaluate the model under varied conditions and expand the training dataset to include a broader range of weather patterns.

ENHANCE MODEL ARCHITECTURE:

- Refine the model to handle diverse data more effectively and consider ensemble methods or cross-validation to improve robustness.

NUANCED EVALUATION:

- Focus on specific weather types and locations to get a clearer picture of the model's strengths and weaknesses, rather than relying on overall accuracy.



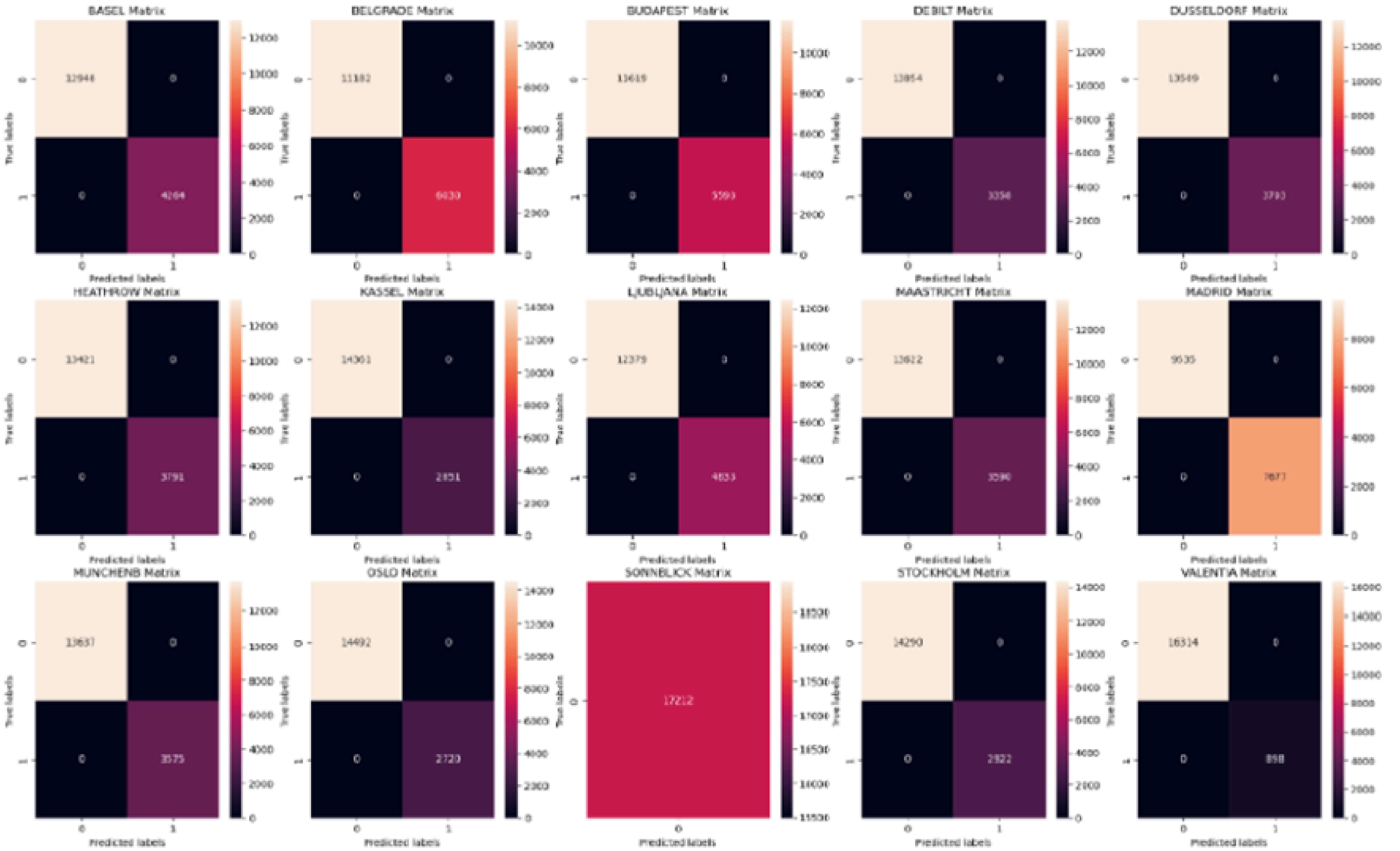
Model Evaluation and Performance Analysis

Decision Trees

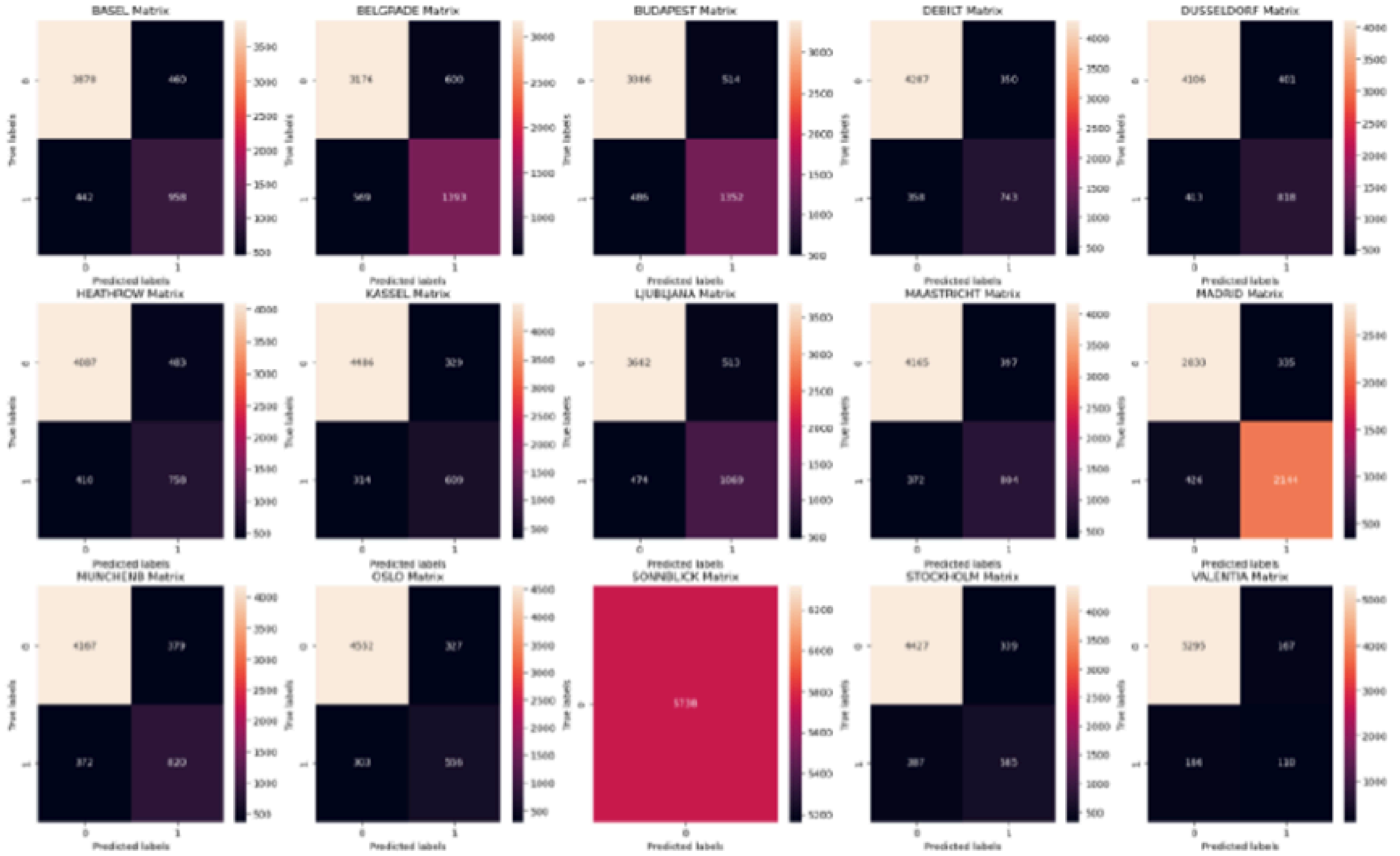
.

Upon evaluation, the decision tree model exhibits an accuracy of 46.1% on the training dataset and 47.5% on the test dataset. These performance metrics suggest that while the model is capable of learning from the training data, it demonstrates limited generalization ability when applied to unseen data. This discrepancy highlights a potential issue of overfitting, where the model excessively captures noise and specifics of the training set, resulting in suboptimal performance on new data.

Train Accuracy Score



Test Accuracy Score



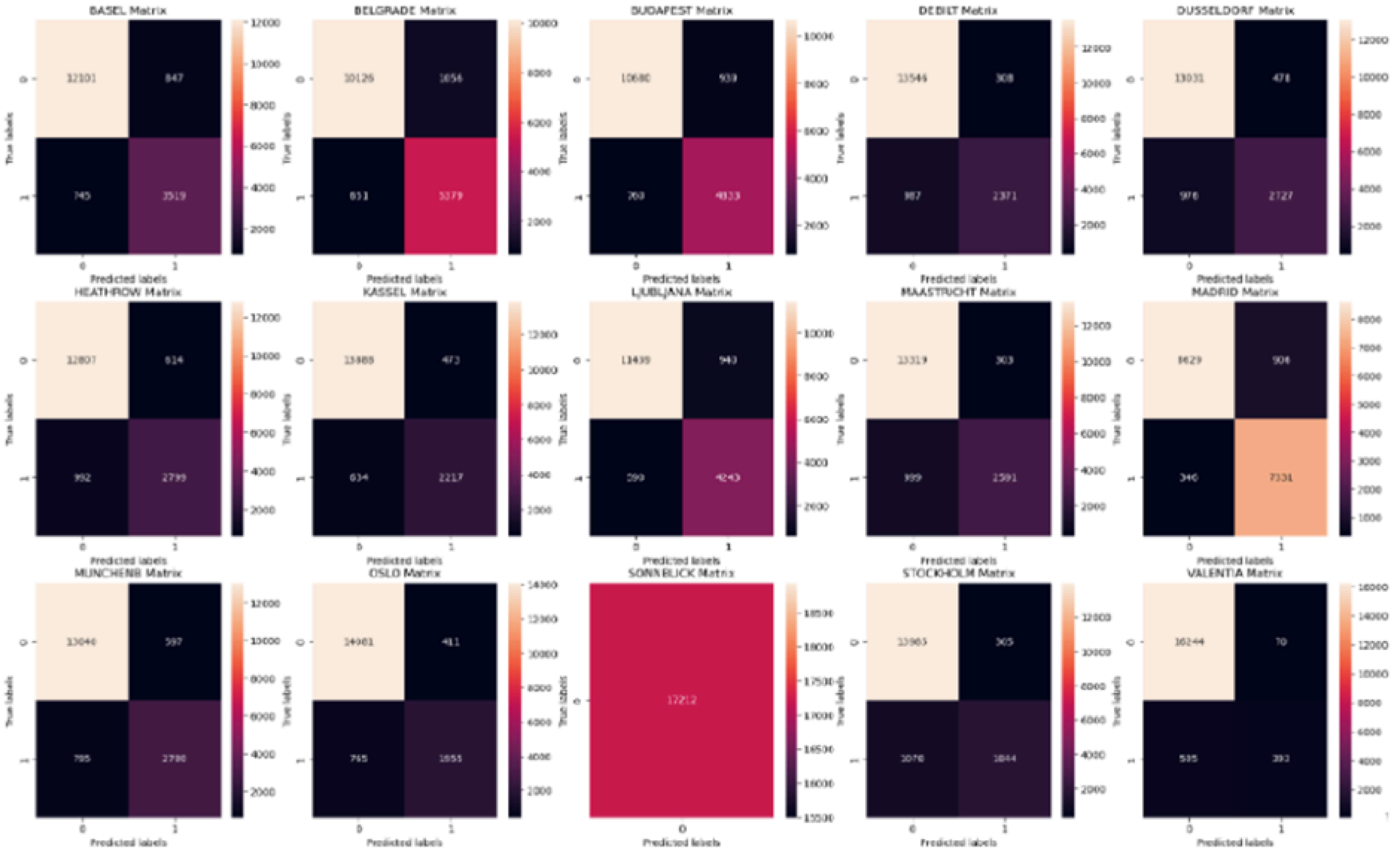
Model Evaluation and Performance Analysis

Upon evaluation, the decision tree model exhibits an accuracy of 46.1% on the training dataset and 47.5% on the test dataset. These performance metrics suggest that while the model is capable of learning from the training data, it demonstrates limited generalization ability when applied to unseen data. This discrepancy highlights a potential issue of overfitting, where the model excessively captures noise and specifics of the training set, resulting in suboptimal performance on new data.

Artificial Neural Networks (ANNs)

Final Training Data

- ANNs will model complex patterns in climate data through interconnected layers of neurons. Training the network with historical data will enable accurate predictions of future weather conditions. This powerful method will assist ClimateWins in anticipating and planning for extreme weather events.

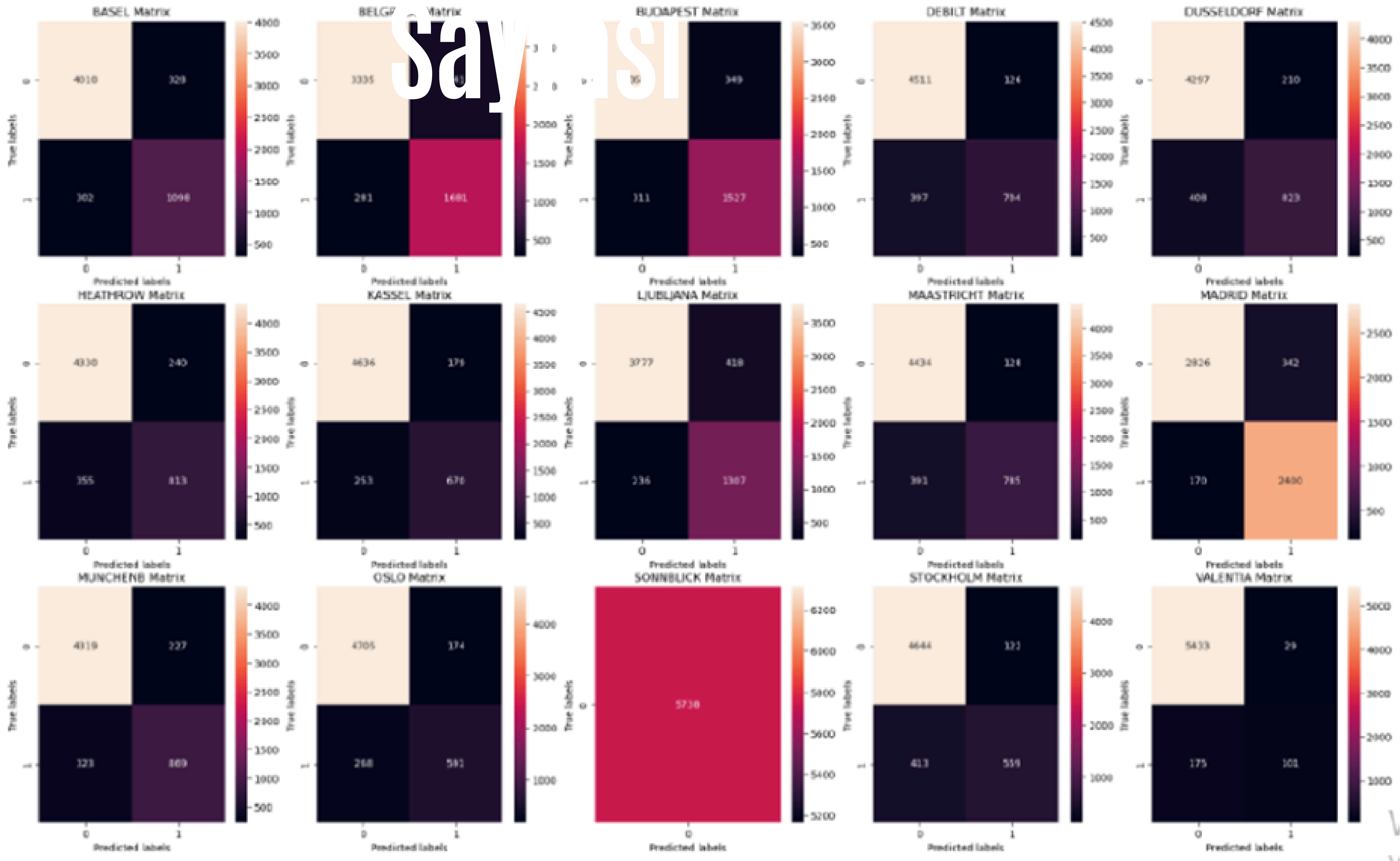


CAN YALÇIN

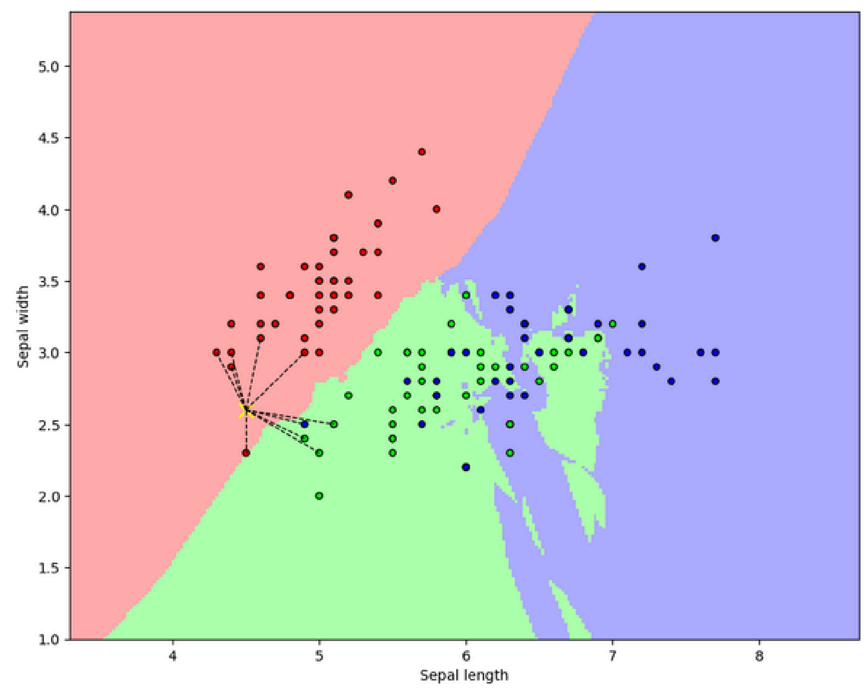
Final testing data

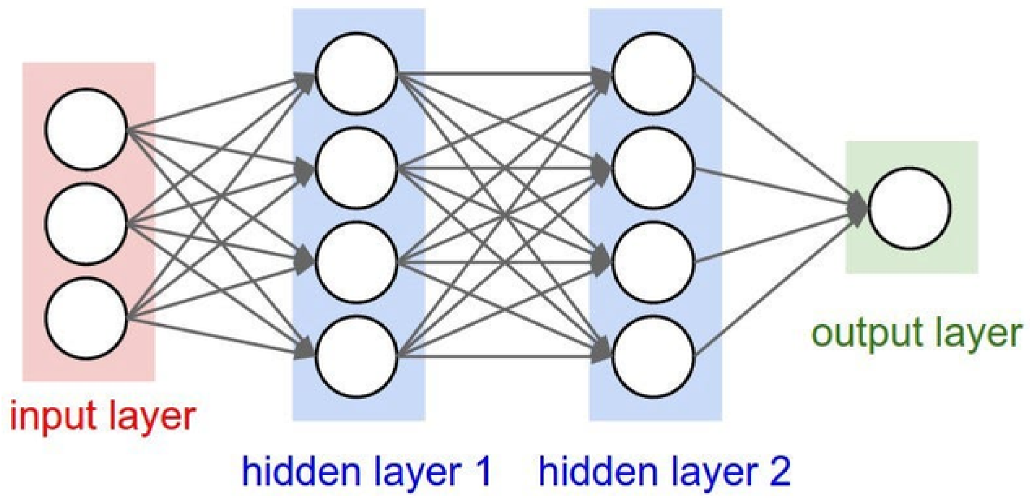
Neural Network Model Evaluation and Recommendation

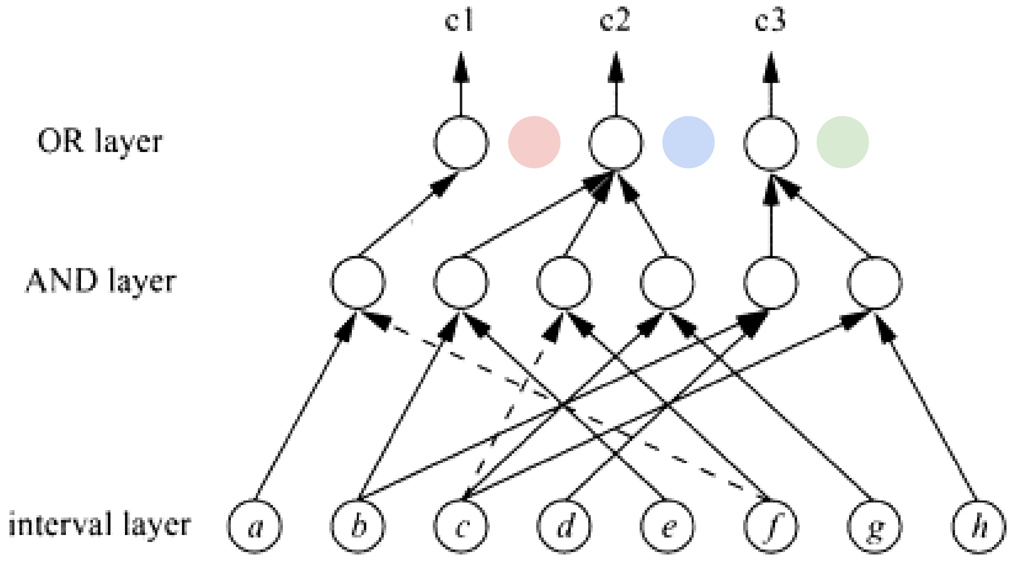
The neural network model is structured with 3 layers, consisting of 70, 60, and 60 nodes respectively, and was trained with a maximum of 1000 iterations and a tolerance level of 0.0003. The model achieved an accuracy of 53.8% on the training data and 49.3% on the testing data. These results indicate that while the neural network can learn from the training data, its ability to generalize to unseen data is limited, suggesting potential overfitting.



CAN YALÇIN







K-NEAREST NEIGHBOURS (KNN)

Accuracy of KNN, Training Set: 88.83%
Accuracy of KNN, Test Set: 86.72%

ARTIFICIAL NEURAL NETWORK (ANN)

Accuracy of ANN, Training Set: 53.8%
Accuracy of ANN, Test Set: 49.3%

DECISION TREE MODEL

Accuracy of DC, Training Set: 46.1 %
Accuracy of DC, Test Set: 47.5%

Conclusion

The KNN model demonstrates the highest accuracy rate, making it the recommended choice for ClimateWins to predict suitable picnic weather. However, the models predicting weather conditions for Sonnblick might be overfitting. Despite achieving a 100% accuracy rate for Sonnblick by consistently predicting unpleasant weather, this could be misleading due to the lack of varied data (pleasant weather conditions) for training. This indicates that the model is not well-rounded and may fail under different circumstances, thereby artificially inflating the overall accuracy.